

CSE190/291

How The Web Tracks You / Targeted Advertising

Course introductions, definitions, administrivia

UC San Diego

Let me address the elephant in the room

Let me address the elephant in the room

- Why does this class have two names...?

Let me address the elephant in the room

- Why does this class have two names....?
- This class is **unofficially co-scheduled** — with CSE 291, an graduate special topics course
 - Most things about these classes are shared
 - Same staff, same PAs, same exams, same Piazza...

Let me address the elephant in the room

- I am not Kristen Vaccaro
- This class is **unofficially co-scheduled** — with CSE 190, an undergraduate special topics course
- Most things about these classes are shared
 - Same staff, same PAs, same exams, same Piazza...
- But some things are not
 - The name (don't ask)
 - Different Canvases, different Gradescope, different enrollment numbers

whoami



Deepak Kumar
Assistant Professor in CSE



undergrad



grad



postdoc

Background on me (research)

- I work primarily on applied computer security research, my research interests are in *sociotechnical security* (computer security + human-computer interaction)
 - Interests primarily in ways that technology + society interact, and where security or safety problems arise (e.g., online harassment, mis/disinformation, AI generated deepfakes, trust on the Internet, etc.)
- Mostly, I'm a data + systems guy.... e.g.,
 - "What does the marketplace for nonconsensual sexual deepfake creation look like?"
 - "How much toxic content there on Reddit, and what can we learn about attack patterns than inform defenses?"
 - "How can we build better defenses for journalists facing online harassment on social media?"

How'd I get into security & privacy?

- Two versions of my story (both are true)
 - I've always been very interested in technology + society, and computer security is a field that by definition gets to impact both of those interests
 - I *wanted* to do computer architecture research in undergrad, but I didn't get a good enough grade. The security group was looking for students and I got an A in the 127-equivalent, so they took me, and the rest is history
- Started in network security, moved my way towards more human-centered work

I also do other things...



Playwright Deepak Kumar, whose "House of India" play will make its world premiere at the Old Globe in 2025. (The Old Globe)

'House of India'

This world premiere play was written by San Diego resident Deepak Kumar, who is both a playwright and a computer science assistant professor at UC San Diego. The play is set in a family-owned Indian restaurant in a Cleveland-area strip mall, where Ananya has reluctantly taken over after her husband's death. She's faced with mounting bills and disagreement between her children over whether to stick to her husband's traditional menu or modernize with a quick-service, fusion concept. "It's about the American question of holding on to our traditions versus assimilating into the mainstream," Edelstein said. It will be directed by Zi Alikhan (Pasadena Playhouse's "Sanctuary City"). *May 10-June 1.*

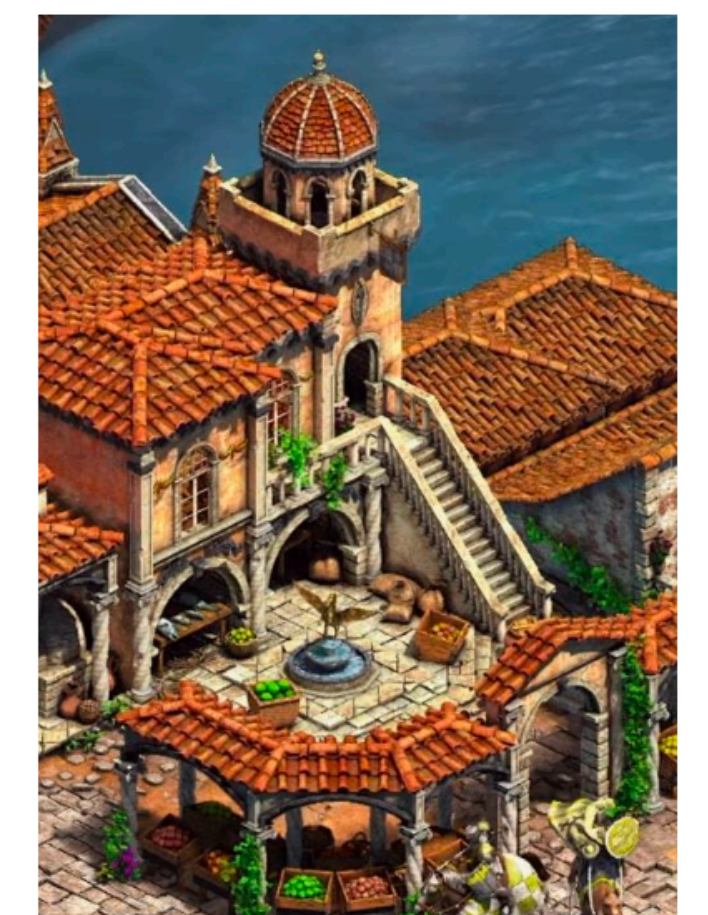


Kristen Vaccaro

Associate Professor

Social media research
content moderation in real-time
political polarization/selective exposure

Auditing/benchmarking AI
dark patterns in LLM front end code
"Auditing AI" MIT Press ← free digital copy



TAs



Haodi Zou

TA

h4zou@ucsd.edu



Edward Zhou

TA

edzhou@ucsd.edu

We want to know about you!

CSE 190 / CSE 291: How the Web Tracks You

Scan to confirm you're enrolled



<https://webtracking.sysnet.ucsd.edu/day1>

I have something to admit

I have something to admit

- **I am not going to read your responses**

I have something to admit

- I am not going to read your responses
- In fact.... I didn't even store your responses. What I *really* wanted... was your data :)

See Browser

YOU ARE MORE UNIQUE THAN YOU THINK

YOU ARE MORE UNIQUE THAN YOU THINK

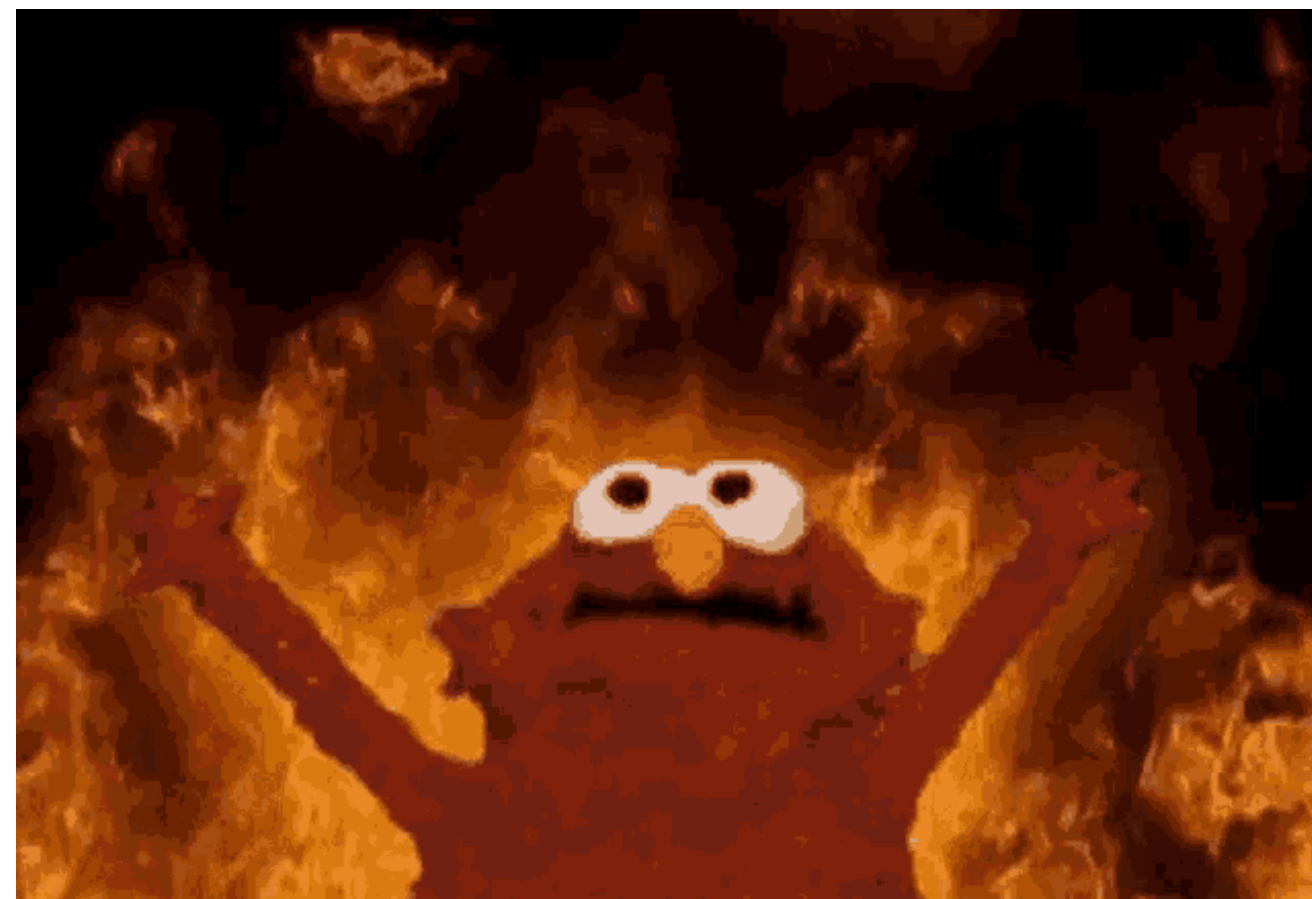
- Your *device*, and how you use your device, is **exceptionally unique to you!**
- Go ahead, try it: <https://tinyurl.com/190me> (or /day1/me)

YOU ARE MORE UNIQUE THAN YOU THINK

- Your *device*, and how you use your device, is **exceptionally unique to you!**
 - Go ahead, try it: <https://tinyurl.com/190me> (or /day1/me)
- Now, you might not have known that... **but the web definitely knows**

YOU ARE MORE UNIQUE THAN YOU THINK

- Your *device*, and how you use your device, is **exceptionally unique to you!**
 - Go ahead, try it: <https://tinyurl.com/190me> (or /day1/me)
- Now, you might not have known that... **but the web definitely knows**
 - And companies are using this information about you **all the time!**



Let's talk about privacy

- Form groups of 2 – 3 with folks around you, and discuss the following questions:
 - What is privacy?
 - Why is privacy important?
 - What is *web tracking*, and how does it relate to privacy?

Let's talk about privacy

- Form groups of 2 – 3 with folks around you, and discuss the following questions:
 - **What is privacy?**
 - Why is privacy important?
 - What is *web tracking*, and how does it relate to privacy?

Definitions: Privacy

- Merriam-Webster online dictionary:
 - The quality or state of being apart from company or observation: **SECLUSION**
 - Freedom from unauthorized intrusion
 - Secrecy
 - "A private matter": **SECRET**

Solove's Taxonomy

- Daniel Solove is a professor at George Washington University Law School
- Considered one of the most influential people in privacy + information technology



Solove's Taxonomy

- Daniel Solove is a professor at George Washington University Law School
 - Considered one of the most influential people in privacy + information technology
- From "A Taxonomy of Privacy," 2006:



"Privacy is a concept in disarray. Nobody can articulate what it means."

Solove's Taxonomy

**Information
Collection**

Surveillance

Interrogation

Solove's Taxonomy

Information Collection

Surveillance

Interrogation

Information Processing

Aggregation

Identification

Insecurity

Secondary use

Exclusion

Solove's Taxonomy

Information Collection

Surveillance

Interrogation

Information Processing

Aggregation

Identification

Insecurity

Secondary use

Exclusion

Information Dissemination

Breach of
Confidentiality

Disclosure

Exposure

Increased Accessibility

Blackmail

Distortion

Appropriation

Solove's Taxonomy

Information Collection

Surveillance

Interrogation

Information Processing

Aggregation

Identification

Insecurity

Secondary use

Exclusion

Information Dissemination

Breach of Confidentiality

Disclosure

Exposure

Increased Accessibility

Blackmail

Distortion

Appropriation

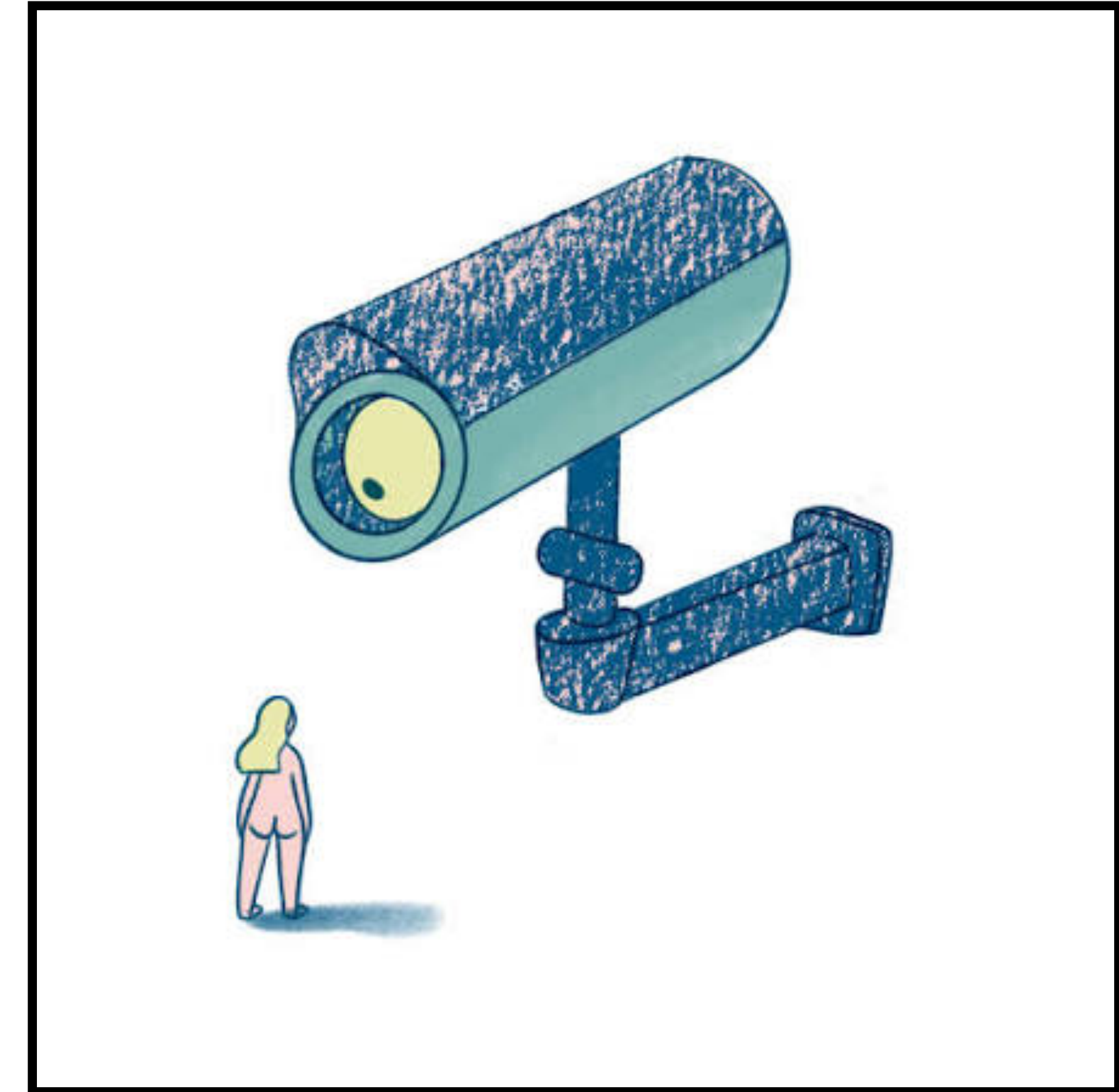
Invasion

Intrusion

Decisional Interference

Privacy is many things...

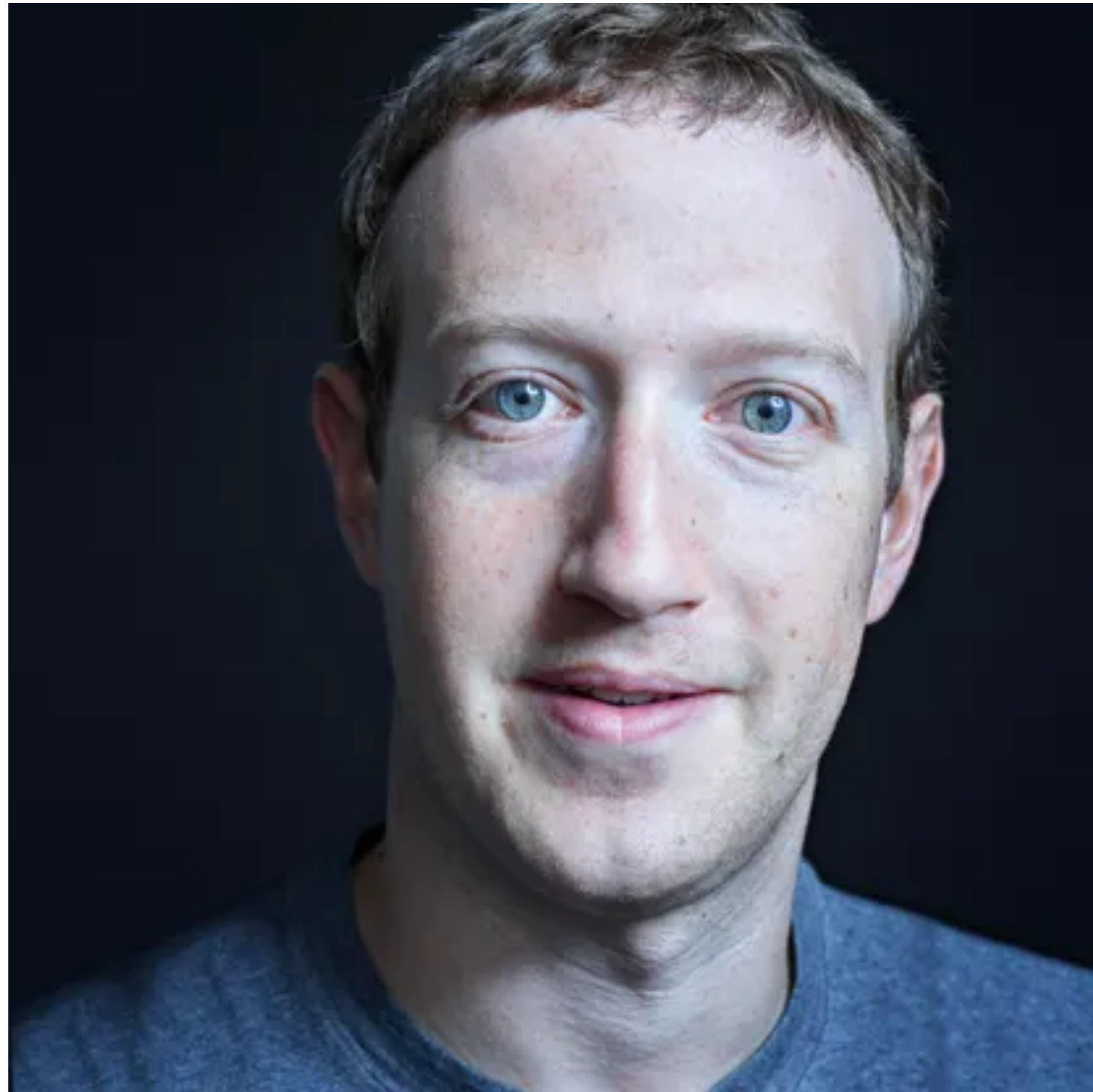
- But ultimately, privacy is an **essential human right**
- More than 180 national constitutions encode a right to privacy!
- <https://www.constituteproject.org/constitutions?key=privacy>



Article 12, Declaration of Human Rights, UN

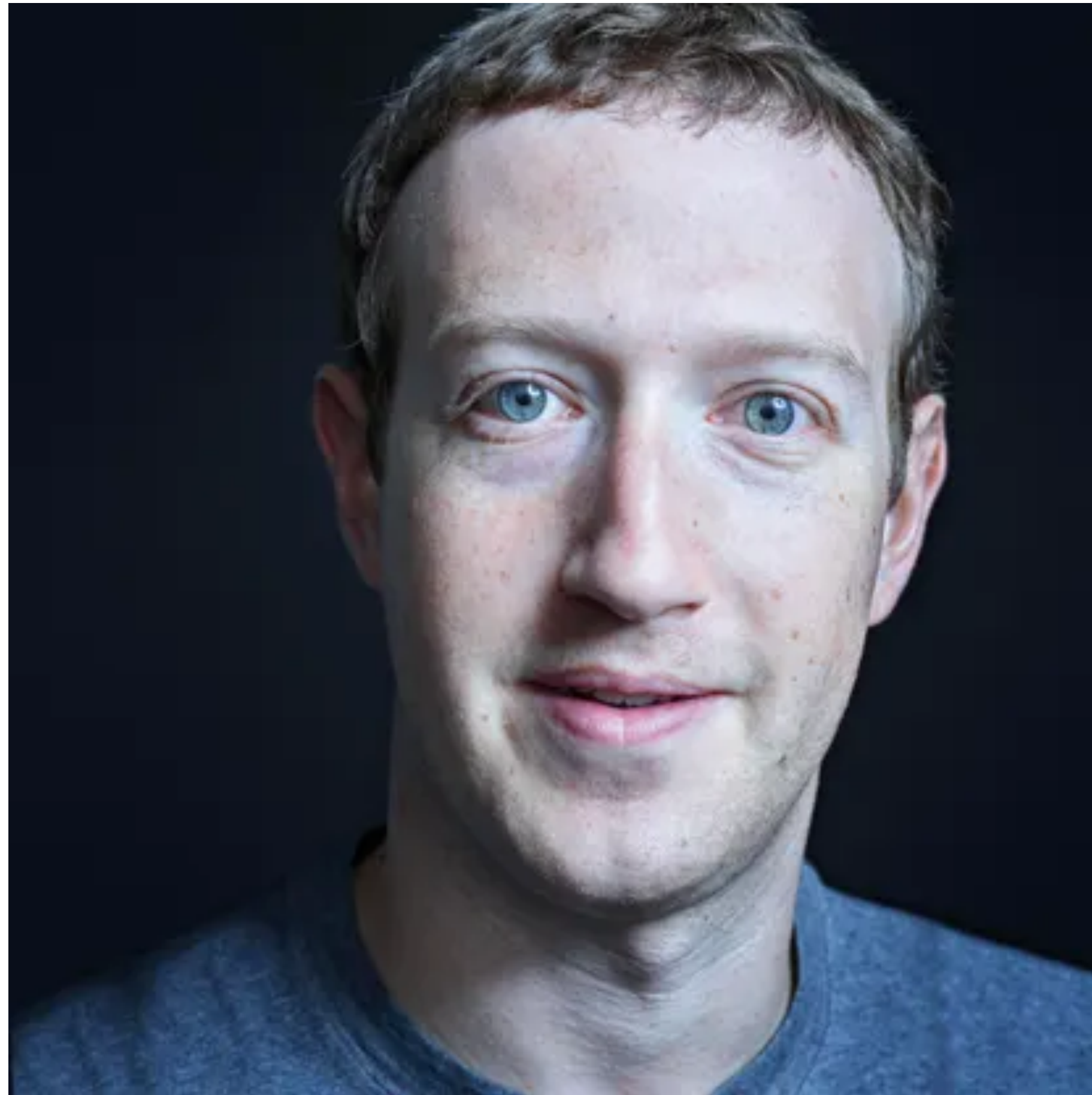
No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.

OK... so... surely tech leaders value our privacy, right?



Mark Zuckerberg, notorious good guy

OK... so... surely tech leaders value our privacy, right?



ZUCK: yea so if you ever need info about anyone at harvard

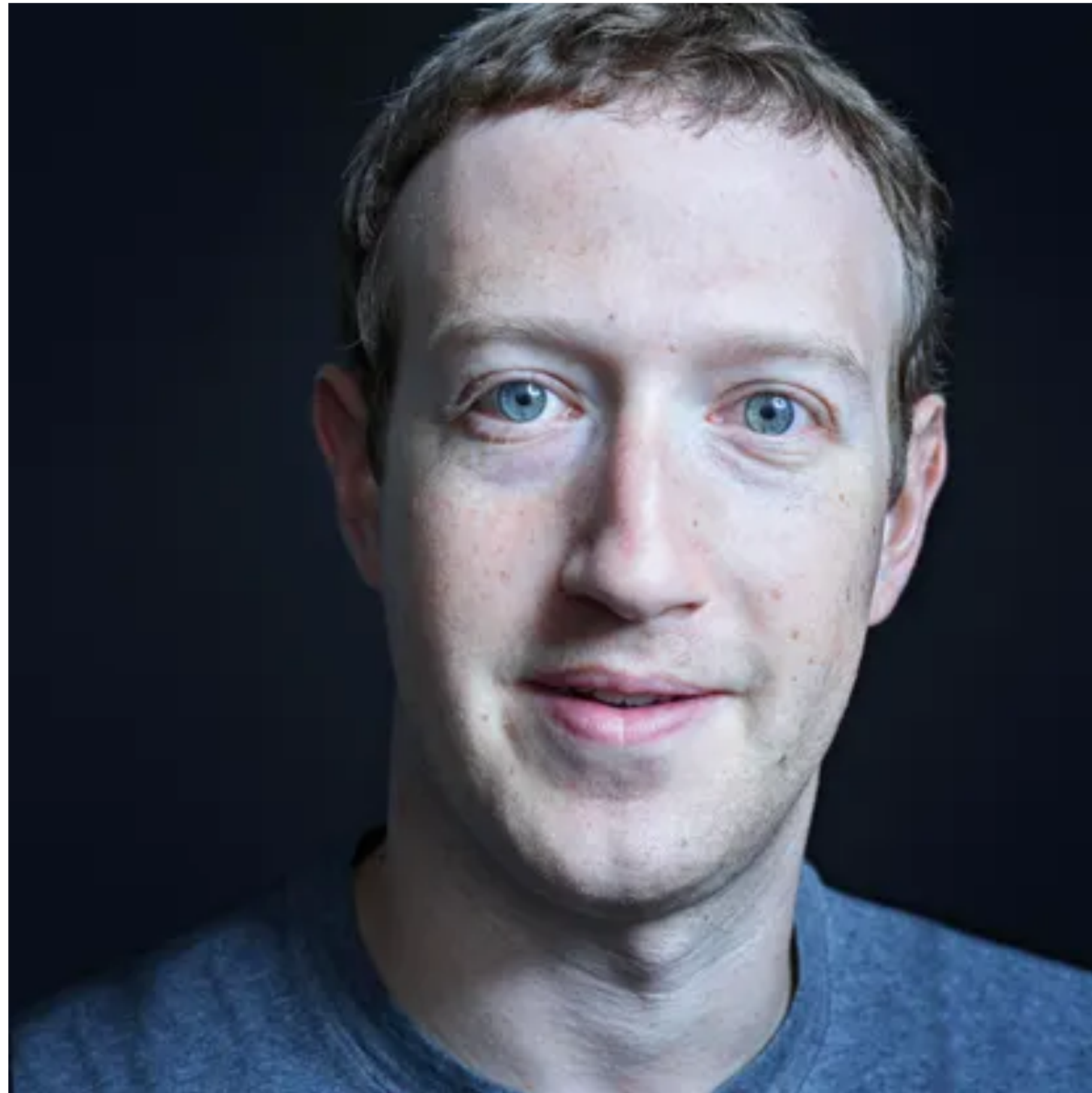
ZUCK: just ask

ZUCK: i have over 4000 emails, pictures, addresses, sns

FRIEND: what!? how'd you manage that one?

Mark Zuckerberg, notorious good guy

OK... so... surely tech leaders value our privacy, right?



ZUCK: yea so if you ever need info about anyone at harvard

ZUCK: just ask

ZUCK: i have over 4000 emails, pictures, addresses, sns

FRIEND: what!? how'd you manage that one?

ZUCK: people just submitted it

ZUCK: i don't know why

ZUCK: they "trust me"

ZUCK: dumb fucks

Mark Zuckerberg, notorious good guy

Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach

Whistleblower describes how firm linked to former Trump adviser Steve Bannon compiled user data to target American voters

- [‘I made Steve Bannon’s psychological warfare tool’: meet the data war whistleblower](#)
- [Mark Zuckerberg breaks silence on Cambridge Analytica](#)

Mark Zuckerberg, notorious good guy

Tech Leaders Demonizing Privacy is Nothing New



Scott McNealy
Sun Microsystems
1999

Tech Leaders Demonizing Privacy is Nothing New



"You have no privacy.
Get over it."

Scott McNealy
Sun Microsystems
1999

Tech Leaders Demonizing Privacy is Nothing New



Scott McNealy
Sun Microsystems
1999



Eric Schmidt
CEO Google
2009

Tech Leaders Demonizing Privacy is Nothing New



Scott McNealy
Sun Microsystems
1999



Eric Schmidt
CEO Google
2009

"If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place."

Tech Leaders Demonizing Privacy is Nothing New



Scott McNealy
Sun Microsystems
1999



Eric Schmidt
CEO Google
2009



Larry Page
Co-Founder Google
2014

Tech Leaders Demonizing Privacy is Nothing New



"We get so worried about these things [privacy] that we don't get the benefits.... Right now we don't data-mine healthcare data. If we did we'd probably save 100,000 lives next year."

Scott McNealy
Sun Microsystems
1999

Eric Schmidt
CEO Google
2009



Larry Page
Co-Founder Google
2014

Tech Leaders Demonizing Privacy is Nothing New



Google's 'Project Nightingale' Gathers Personal Health Data on Millions of Americans

Search giant is amassing health records from Ascension facilities in 21 states; patients not yet informed

By Rob Copeland

Updated Nov. 11, 2019 4:27 pm ET

Scott McNealy
Sun Microsystems
1999

Eric Schmidt
CEO Google
2009

Larry Page
Co-Founder Google
2014

Tech Leaders Demonizing Privacy is Nothing New



Google's 'Project Nightingale' Gathers Personal Health Data on Millions of Americans

Search giant is amassing health records from Ascension facilities in 21 states; patients not yet informed

By Rob Copeland

Updated Nov. 11, 2019 4:27 pm ET

Scott McNealy
Sun Microsystems
1999

Eric Schmidt
CEO Google
2009

Larry Page
Co-Founder Google
2014

This was probably illegal, btw

Tech Leaders Demonizing Privacy is Nothing New



Scott McNealy
Sun Microsystems
1999



Eric Schmidt
CEO Google
2009



Larry Page
Co-Founder Google
2014



Larry Ellison
Co-Founder Oracle
2024

Tech Leaders Demonizing Privacy is Nothing New



Scott McNealy
Sun Microsystems
1999



Eric Schmidt
CEO Google
2009

"Citizens will be on their best behavior, because we're constantly recording and reporting everything that is going on."



Larry Ellison
Co-Founder Oracle
2024

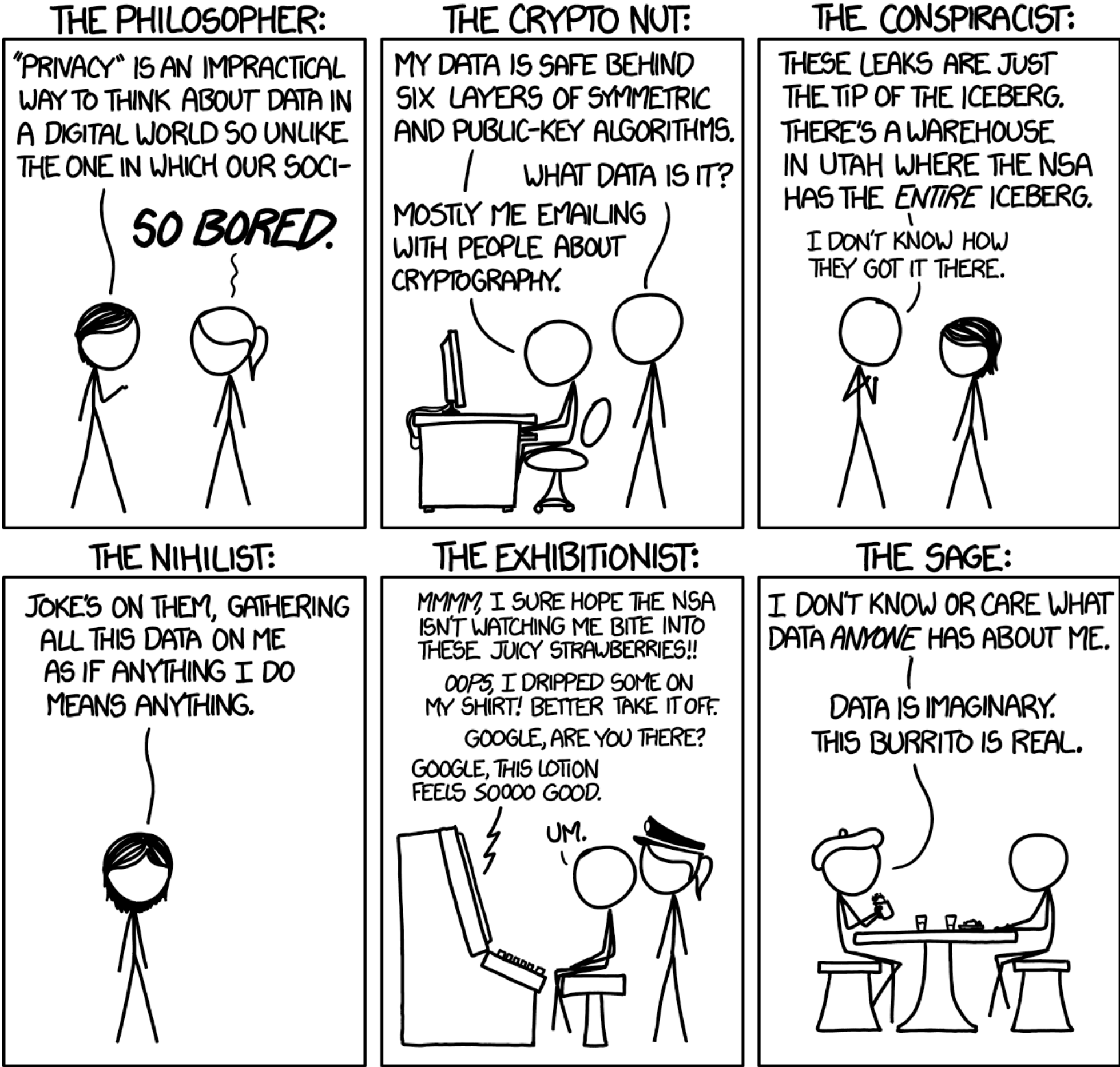
Larry Page
Co-Founder Google
2014

Let's talk about privacy

- Form groups of 2 – 3 with folks around you, and discuss the following questions:
 - What is privacy?
 - **Why is privacy important?**
 - What is *web tracking*, and how does it relate to privacy?

Privacy is highly discussed

OPINIONS ON INTERNET PRIVACY



Majority of Americans feel as if they have little control over data collected about them by companies and the government

% of U.S. adults who say ...

		Companies	The government
Lack of control	They have very little/no control over the data __ collect(s)	81%	84%
Risks outweigh benefits	Potential risks of __ collecting data about them outweigh the benefits	81%	66%
Concern over data use	They are very/somewhat concerned about how __ use(s) the data collected	79%	64%
Lack of understanding about data use	They have very little/no understanding about what __ do/does with the data collected	59%	78%

Note: Those who did not give an answer or who gave other responses are not shown.

Source: Survey conducted June 3-17, 2019.

"Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information"

PEW RESEARCH CENTER

Privacy is valuable

What major tech company markets themselves on privacy?

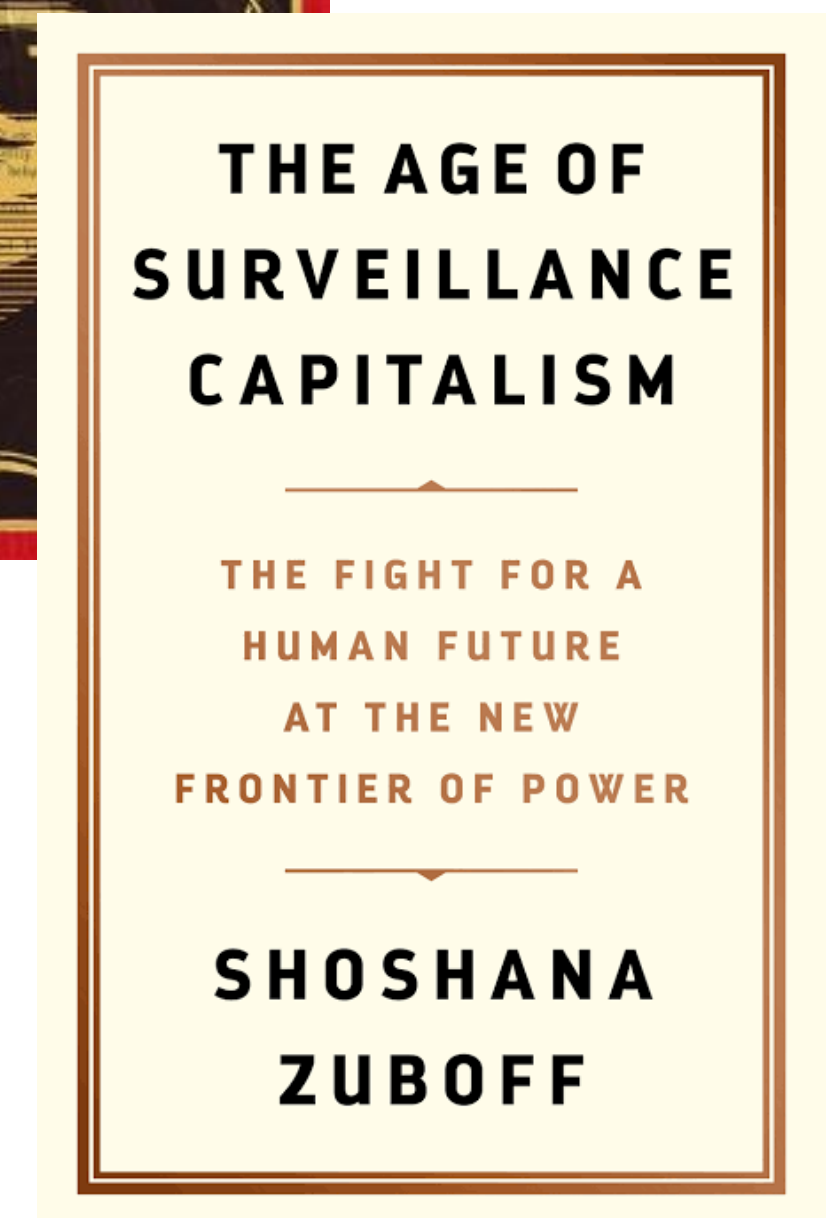
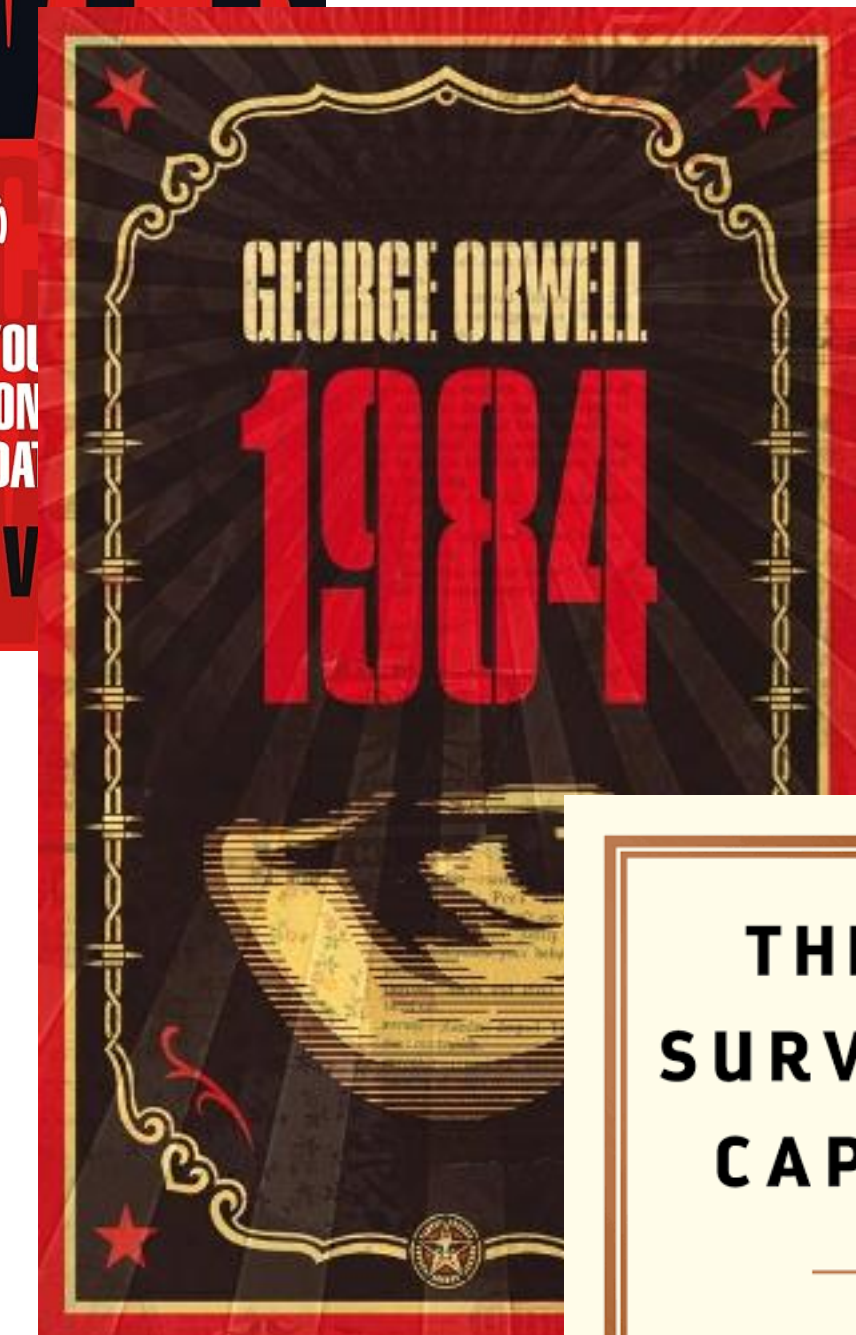
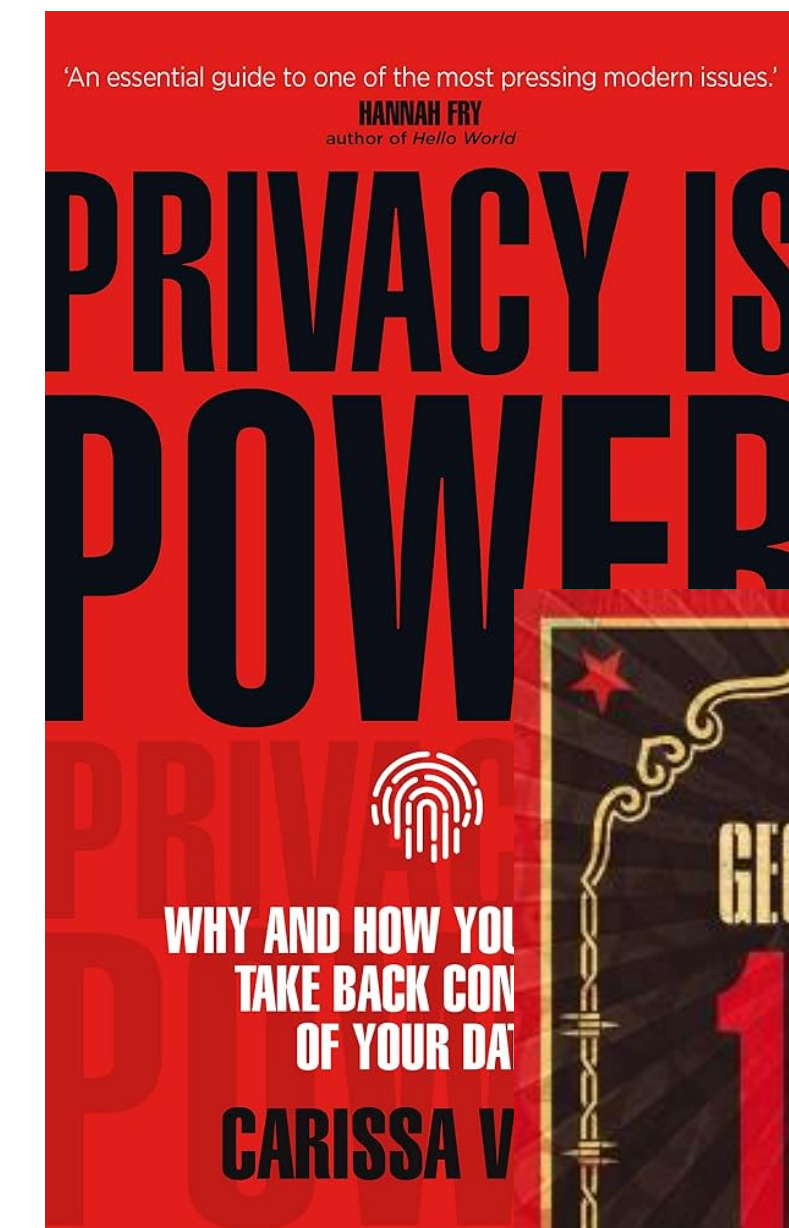
Privacy is valuable

What major tech company markets themselves on privacy?



Privacy is power

- “Control over information is control over people” — Carissa Veliz
- Basic argument is: a lack of privacy inherently gives someone *control* over you
- Zuboff’s formulation: “Who Knows, Who Decides, and Who Decides Who Decides?”
- Poll: Caught in a private moment? Were you thrilled about that?



Privacy is freedom



What does web tracking have to do with any of this?

What does web tracking have to do with any of this?

- Today, your privacy is an explicit tradeoff being made in a machinery of tracking and surveillance, entirely powered by the Internet
 - Why?

What does web tracking have to do with any of this?

- Today, your privacy is an explicit tradeoff being made in a machinery of tracking and surveillance, entirely powered by the Internet
 - Why?
 - Largely... advertising.

What does web tracking have to do with any of this?

- Poll: How many people believe your phone is listening to you and using that data to sell you advertisements?

What does web tracking have to do with any of this?

- Poll: How many people believe your phone is listening to you and using that data to sell you advertisements?

Good news!

Why phones that secretly listen to us are a myth

5 September 2019

Joe Tidy

Cyber security reporter, BBC News

Share

Save

Add as preferred on Google

Is your phone listening to you? Yeah, but probably not to target ads.

The specifics of the theory that your phone is eavesdropping probably true, but the paranoia is justified.

Updated January 7, 2025 More than 1 year ago

6 min 202

Make us preferred

No, Your iPhone Isn't Listening to You. But the Truth Is Even Worse

If you want a monster, it's not a hot mic. It's more about shadowy presences lurking on the internet.

What does web tracking have to do with any of this?

- Poll: How many people believe your phone is listening to you and using that data to sell you advertisements?

Bad news: the tracking is really *that* good.

This class: Web Tracking, Advertising, and You

- This class is about **how the tracking ecosystem works, technically, under the hood**
- You will get hands-on experience with the technical tools that are used to track you
- You will develop an appreciation for why tracking exists, its history, and the ethical and its societal ramifications
- You will understand, at least a bit more, about what you can do about it

My take on privacy and web tracking

- Advertising has evolved as the de-facto “price we pay” for a free-ish Internet
 - It’s very interesting to learn how we got here, and what it means for where we’re going
- Privacy is an important, evergreen area — and technology is playing a role
 - As computer scientists, it is our **obligation** to understand how *the things we build* interact with the world around us (in good + bad ways)
- Web tracking and advertising is the specific implementation of our current privacy battle, but privacy battles are older than the web (and will live beyond the web too)

Course Ethos + Logistics

Course info

- Website: <https://kumarde.com/webtracking-fa26/>
 - Slides posted right before class so you can follow along (not today, obviously)
- Two versions of...
 - Canvas
 - Gradescope
- Piazza for communication (avoid email if you can)
- Office Hours
 - See the course webpage for most up-to-date office hours

This is the first iteration of this class ever.

- **EVERYTHING IS NEW!**
 - All the lectures are new
 - All the assignments, infrastructure, grading is new
 - All the exams are new



This is the first iteration of this class ever.

- **EVERYTHING IS NEW!**
 - All the lectures are new
 - All the assignments, infrastructure, grading is new
 - All the exams are new
- Things will **for sure** break during the quarter.
Thanks for your patience.



This is the first iteration of this class ever.

- **EVERYTHING IS NEW!**
 - All the lectures are new
 - All the assignments, infrastructure, grading is new
 - All the exams are new
- Things will **for sure** break during the quarter.
Thanks for your patience.
- There is no other class like this offered anywhere — where people get hands on experience doing this kind of work!



Goals and non-goals of this course

This is a hands-on course

- **Goals**

- Provide you with a hands-on foundation of web, web privacy, web tracking concepts, applied to concrete topics areas and hands-on PAs
 - Stateful tracking — cookies, inference, real time bidding
 - Stateless tracking — fingerprinting, uniqueness
 - Ad blocking — what does it take to build a robust ad blocker today?
- Understanding how things work, how they break, and how to fix them

- **Non-goals**

- A deep dive into the other important aspects of privacy — policy, regulation, activism, etc.
- We can't cover every mechanism used to track you (but you can study them after this class)

This course, broadly

Topic areas

- We'll cover the following topic areas as best we can in ten weeks:
 - Web & Databases
 - Cookie-based stateful tracking
 - Machine learning + inference for serving ads
 - Online advertising, real-time bidding
 - Stateless tracking (how do I track you without a cookie?)
 - Ad blocking
 - Surveillance capitalism and the ethics of privacy

Course materials

- **No textbook** for the class, but here's some books I recommend anyway if you're curious...
 - Privacy in Context, Helen Nissenbaum (proper textbook)
 - Surveillance Capitalism, Shoshana Zuboff
- Other readings, we'll provide as we go on the course website for additional context and knowledge

Structure + Grading of this course

- Attendance (5% of your grade)
- Programming Assignments (40% of your grade)
- Midterm (20% of your grade)
- Final (35% of your grade)

Attendance

- Attendance is worth 5% of your grade
- We will do an attendance check-in once per lecture — at some arbitrary time (could be in the beginning, right at the end, etc.)
- You can attend either of the classes on any given day, attendance is all counted together

Programming Assignments

- You'll get hands-on experience with web tracking in the following domains:
 - **Web & Databases** (PA1, parts 1 + 2)
 - **ML Inference + Real-time Advertising** (PA1, parts 3 + 4)
 - **Stateless Tracking** (PA2)
 - **Ad Blocking** (PA3)
- Schedule of release + due dates are already available on website
 - *This class moves very fast. There is always an active PA. PA1, 1+2 is already out!*

Programming Assignments

- Programming assignments MAY be done in teams (if you wish); teams do not need to stay the same between PAs
- PA1 is 18% of your grade, PA2 is 12%, and PA3 is 10%
- Students MUST NOT collaborate with anyone outside of their partner; any unsanctioned group work is academic integrity violation and will be reported

What you will need to know for PAs

- In order to accomplish the PAs, you will need to be familiar with (or learn): Python, JavaScript, HTML, SQL, cloud deployments... **at least**
- We'll cover the high levels of these in class (and some specifics, sometimes), but you should be prepared to learn stuff on your own
- Plus, we have lots of staff office hours, so use us as the resources we are when you get stuck!

Rules for PAs

- PAs are due at 11:59:59 PT on the due date listed
- You get two late days on PAs (both teammates need to have a late day to use otherwise you don't get one) — these are applied **automatically**
- Regrades are very much the exception, not the norm
 - We reserve the right to completely regrade your assignments :)
- **NO CHEATING**
 - Read and understand the UCSD policy on academic integrity
 - <https://academicintegrity.ucsd.edu>
 - Not OK to copy, paraphrase, translate, etc. somebody else's work
 - If you not sure if something is cheating, **assume it is until you ask about it**

A word on AI...

- Feel free to use AI to help you with your PAs
 - **BUT USE THEM AT YOUR OWN PERIL**
- You must attest to your usage of AI in Gradescope. Otherwise, it's an academic integrity violation (yes, we are checking)
- Note: Exams (which remember are 55% of your grade) will test both lecture material **and** PA material.
 - If you find a way to use AI to do no metacognition, congrats, but you'll probably do quite poorly on exams (and therefore, do poorly in the course)
 - Plus, you're here at UCSD to learn how to do things on your own...
 - My recommendation? Don't use AI at all, but you're all adults, so up to you

A word on ethics...

- In this class, you will learn how to infringe on the privacy of everyone you know.
 - We learn these because understanding how things work is crucial to building **defenses** against those infringements
- You are obligated to use this knowledge *ethically*
 - You **MUST NOT** attack others... and in addition to being unethical, you have **no idea** what laws you're breaking if you do so (e.g., HIPAA, FERPA, etc.)

Exams

- Two exams, a midterm (week 6, in class) and a final (during finals week)
 - Midterm 20%, Final 35%
- Structure of both is very similar
 - Litany of multiple choice questions primarily focused on lecture
 - Broader “PA”-style questions that build on what you did for the PA but with different inputs / outputs
- Final is comprehensive but will trend towards the back half of class (PA2 + PA3)
- **If you do better on the final than the midterm, your midterm grade will automatically be replaced with your final grade.**

Course Vibes

Community-centric learning

- The classroom is *community*
 - Get to know one another! The course is only made better when you know each other
 - Come ready to probe with questions, comments, concerns, thoughts, etc.
 - Discussions should be *respectful*, understanding everyone is here to contribute and to learn

Immediate Logistics

- **PA1 parts 1 + 2 are released on course webpage!** It is the basics of web programming – servers, clients, databases, and deployment. You will build on these in parts 3 + 4.
 - Part 1, due on 10/1 at 11:59:59pm
 - Part 2, due on 10/13 at 11:59:59pm
 - Find a teammate, get to work! Piazza has the teammate search feature open should you need to coordinate.
 - Submission on Gradescope via Canvas
- Complete the #FinAid survey on Canvas (necessary to support those w/ financial aid needs) — included on respective Canvases

Next time...

- We'll talk about the web, clients, servers, protocols, HTML.... and all that jazz
- PA1 is out! Good luck!